IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121238

ML-Driven Spam Classification Model

Nandini P Gowda.[1], Jnanashree TR.[2], N Govind Prasad.[3], Vibha Datta[4]

Associate Professor, Department of CS&D, K. S. Institute of Technology, Bengaluru, India[1]

Student, Department of CS&D, K. S. Institute of Technology, Bengaluru, India[2]

Student, Department of CS&D, K. S. Institute of Technology, Bengaluru, India[3]

Student, Department of CS&D, K. S. Institute of Technology, Bengaluru, India[4]

Abstract: The rapid expansion of digital communication has significantly increased the amount of spam across platforms such as SMS, email, URLs, and social media. These spam messages often contain phishing links, fake offers, and harmful advertisements that threaten user security. To address this issue, the project proposes a Machine Learning–based Spam Classification System that can automatically detect and filter spam from various communication sources. The system begins by cleaning and preprocessing text data through steps like tokenization, stop-word removal, and normalization. It then uses TF-IDF to convert textual information into numerical features. Multiple machine learning models are trained to accurately distinguish between spam and legitimate messages. The system learns underlying patterns that help identify spam more effectively. Its performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The proposed solution minimizes false detections and improves reliability. It is capable of processing large volumes of data and can be easily integrated into real-time applications. Overall, the system strengthens communication security and builds user trust.

Keywords: Spam Classification, Machine Learning, SMS Spam Detection, Email Spam Filtering, URL Analysis, Social Media Spam, Text Classification, TF-IDF, Cyber Security.

I. INTRODUCTION

The growing use of digital communication platforms such as SMS, email, URLs, and social media has led to a noticeable rise in spam messages. These unwanted messages not only create inconvenience for users but also introduce serious security risks, including phishing scams, fraud, and the spread of malware. Traditional rule-based spam filters often fail to keep up with the constantly changing techniques used by spammers. To overcome these limitations, machine learning offers a smarter way to automatically identify spam by learning from data patterns. This project focuses on building a spam classification system using machine learning techniques. The system examines text content to differentiate between spam and genuine messages. Effective data preprocessing and feature extraction help improve classification accuracy. The model works across multiple communication platforms and adapts to emerging spam trends. It also aims to minimize false detections. Overall, the project contributes to safer and more reliable digital communication.

II. LITERATURE REVIEW

Recent research highlights the increasing need for effective spam detection systems due to the rapid growth of digital communication platforms such as SMS, email, URLs, and social media. Traditional rule-based spam filters struggle to adapt to evolving spam patterns and sophisticated phishing techniques. As a result, machine learning-based approaches have gained significant attention for their ability to learn from data and improve detection accuracy over time. Several studies have explored different algorithms, feature extraction techniques, and datasets to enhance spam classification performance while reducing false positives.

Specific methodologies and findings from previous research include:

i. Text-Based Spam Classification:

Early studies focused on detecting spam in SMS and email using text-based features. Techniques such as tokenization, stop-word removal, and bag-of-words models were widely used. Machine learning algorithms like Naïve Bayes and Support Vector Machines demonstrated reliable performance due to their simplicity and efficiency in handling textual data.

ii. URL and Phishing Detection:

Research on URL spam detection emphasized analyzing link structure, domain features, and suspicious keywords.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121238

Studies showed that combining textual content with URL-based features significantly improves phishing detection accuracy. Random Forest and Logistic Regression models performed well in identifying malicious links.

iii. Social Media Spam Detection:

With the rise of social media platforms, researchers investigated spam detection using user behavior, message frequency, and content similarity. Machine learning models trained on social media data effectively identified promotional spam, fake profiles, and scam messages. Feature engineering played a crucial role in improving classification results.

iv. Advanced Machine Learning Techniques:

Recent works explored ensemble models and hybrid approaches combining multiple classifiers to enhance detection accuracy. Feature extraction methods such as TF-IDF and n-grams were found to improve model performance. These studies concluded that machine learning-based spam classification systems are scalable, adaptable, and suitable for real-time applications.

III. PROBLEM STATEMENT

Many researchers have studied the problem of spam in digital communication, as spam messages continue to increase across SMS, email, URLs, and social media platforms. Earlier spam detection methods mainly depended on fixed rules and keyword matching, which were not effective against changing spam techniques. To overcome these limitations, machine learning approaches have been widely introduced. Algorithms such as Naïve Bayes, Support Vector Machines, and Decision Trees have shown promising results in identifying spam messages. Studies highlight the importance of preprocessing steps like text cleaning, tokenization, and stop-word removal. Feature extraction techniques such as TF-IDF help convert text into meaningful numerical data. Researchers also found that combining multiple features improves accuracy. URL analysis further strengthens phishing detection. Social media spam studies include behavior-based analysis. Overall, machine learning has proven to be a reliable and scalable solution for spam detection.

IV. METHODOLOGY

A. *Objectives*

The primary objective of this project is to develop an efficient spam classification system that can accurately identify spam across multiple communication platforms.

i. Spam Detection Accuracy: To effectively distinguish between spam and legitimate messages using machine models.
ii. Multi-Platform Support: To detect spam in SMS, email, URLs, and social media messages using a unified approach.
iii. Security Enhancement: To reduce exposure to phishing, fraud, and malicious content.
iv. Reduced False Positives: To minimize incorrect classification of legitimate messages as spam.
v. Scalability: To design a system capable of handling large volumes of data efficiently.

B. *Design Methodology*

The project follows a structured machine learning-based design approach, consisting of sequential and iterative stages to ensure accurate and reliable spam detection. The methodology includes data collection, preprocessing, feature extraction, model training, and evaluation.

The general steps followed include:

i. i. Data Collection: Datasets containing spam and non-spam messages from SMS, email, URLs, and social media sources are gathered.
ii. Data Preprocessing: Text data is cleaned by removing unwanted characters, stop-words, punctuation, and converting text into a standardized format.
iii. Feature Extraction: TF-IDF is used to transform textual data into numerical vectors suitable for machine learning models.
iv. Model Training: Multiple machine learning algorithms are trained to learn patterns that distinguish spam from legitimate messages.
v. Evaluation and Optimization: Models are evaluated using accuracy, precision, recall, and F1-score to select the most effective classifier.

C. *System Architecture and Requirements*

The proposed spam classification system follows a **multi-layered architecture** designed to efficiently process messages from different communication platforms and accurately classify them as spam or legitimate.

i. Input Layer:
Collects messages from multiple communication sources for analysis.
ii. Preprocessing Layer:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121238

Cleans text data and extracts features using TF-IDF.

iii. Classification Layer:

Applies machine learning models to detect spam patterns.

iv. Output Layer:

Displays results as spam or legitimate messages.

The system must support accurate multi-platform spam detection, fast processing, scalability, and secure data handling. It should minimize false positives and adapt to evolving spam patterns.

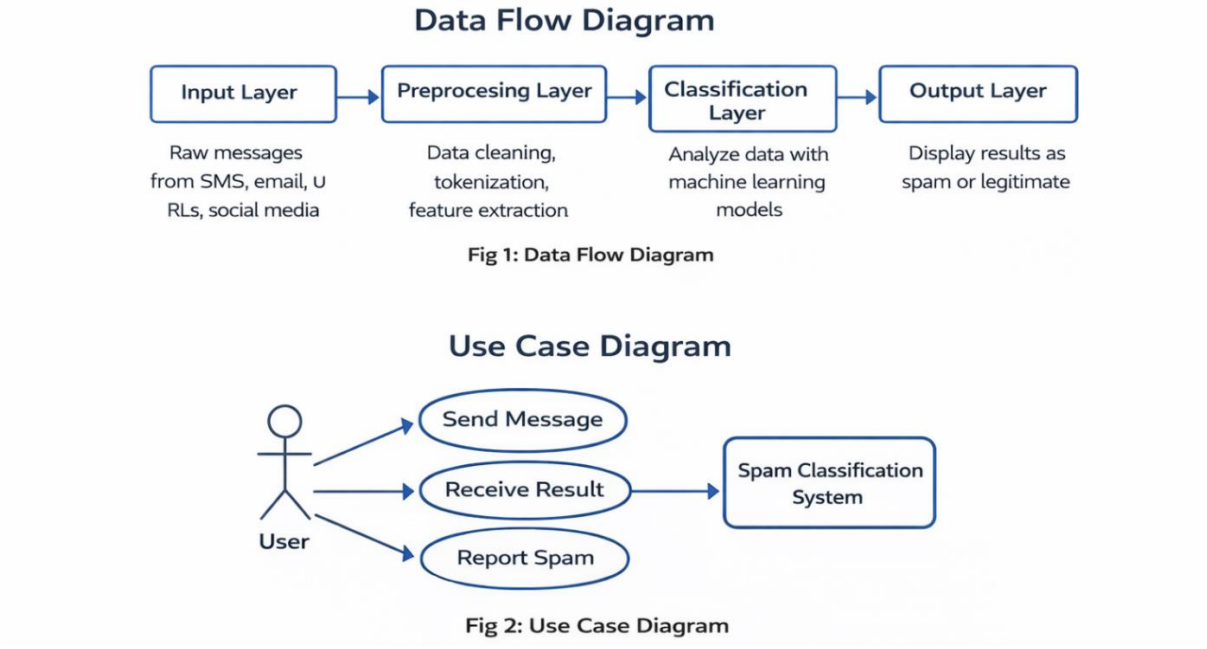

Fig 1: Data Flow Diagram

Fig 2: Use Case Diagram

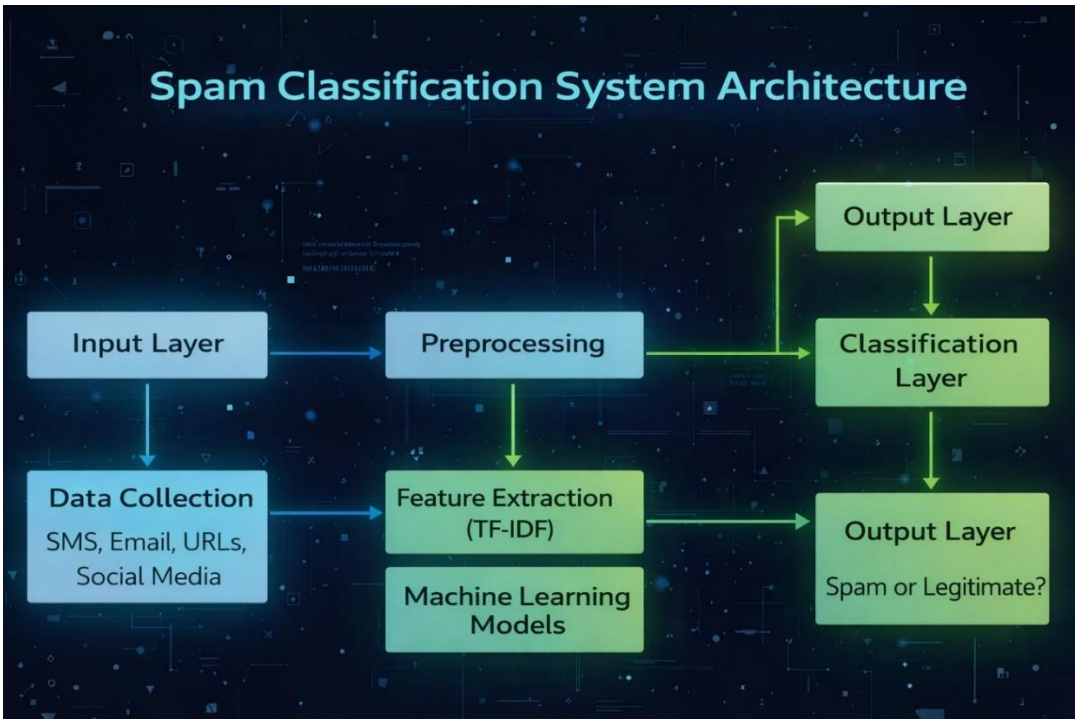

Fig 3: System Architecture

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121238

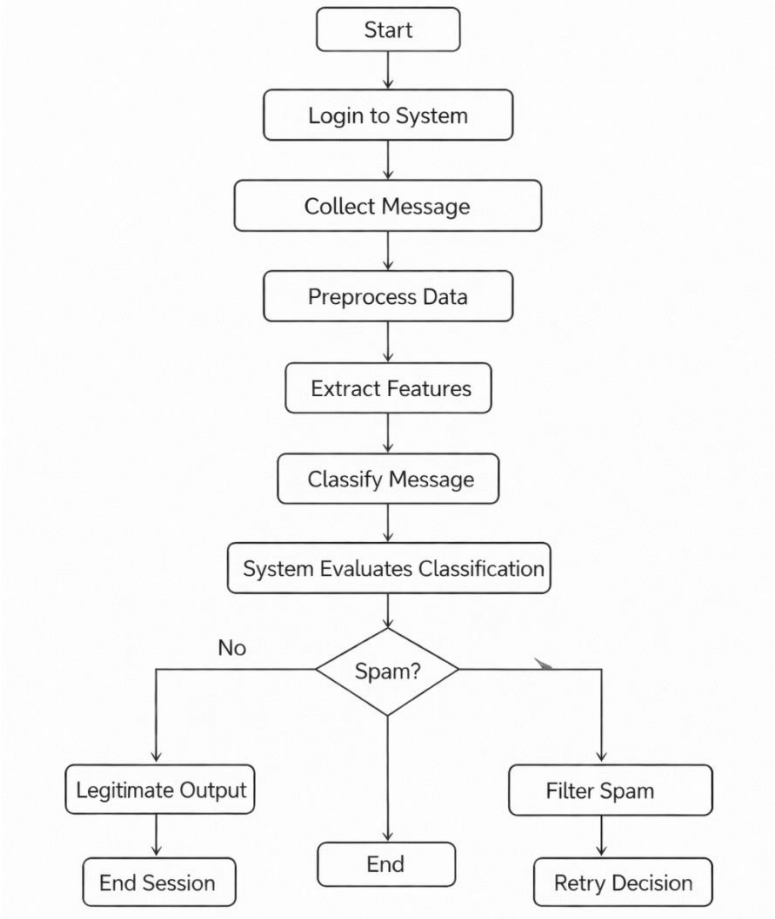

Fig 4: Activity Diagram

D. *Hardware and Software Requirements*

i. Hardware Requirements include A PC or laptop with a minimum of an Intel Core i3 processor, 8 GB RAM, 256 GB storage, and a stable internet connection is required.

ii. Software Requirements include The system uses Windows/Linux, Python, and machine learning libraries such as NumPy, Pandas, and Scikit-learn. Development is carried out using tools like Jupyter Notebook or VS Code.

V. RESULTS AND DISCUSSION

The testing of the spam classification system showed encouraging results across SMS, email, URLs, and social media messages. After cleaning and processing the data, the machine learning models were able to clearly understand the difference between spam and genuine messages.

A. *Spam Detection Performance*

The system performed well in identifying spam messages with good accuracy. Most spam content was correctly detected, while genuine messages were rarely misclassified. This shows that the feature extraction and learning process worked effectively.

B. *System Reliability and Adaptability*

The model was able to handle different types of spam and adapt to new message patterns. Since it learns from data, the system can improve over time and remain effective even as spam techniques change across platforms.

C. *Challenges and Observations*

A few errors occurred when messages were very short or unclear, making them harder to classify. However, these cases were limited, and the overall system remained reliable. With more data and improved techniques, the accuracy can be further enhanced.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121238

VI. CONCLUSION

In today's digital world, spam messages have become a common problem across SMS, emails, social media platforms, and online links, often causing inconvenience and security risks to users. This project focused on building a machine learning–based spam classification system that can automatically identify and filter spam messages from genuine content. By using text preprocessing techniques and machine learning algorithms, the system learns patterns from past data and makes accurate predictions on new messages. The results show that machine learning provides a reliable and efficient solution compared to traditional rule-based filtering methods. The model helps reduce exposure to unwanted advertisements, phishing attempts, and malicious links, thereby improving user safety and communication quality. Although some challenges remain, such as handling very short or unclear messages, the overall performance of the system is promising. With further improvements like larger datasets and advanced learning models, the system can become even more accurate and adaptable. Overall, this project demonstrates how machine learning can play a significant role in creating safer and cleaner digital communication environments for users across multiple platforms.

REFERENCES

[1]. Reddy, K.R. and Joshi, G., 2024, December. Innovative Development of Sophisticated Text Mining Architectures for Precision Spam Detection Leveraging NLP Techniques, Neural Networks, and Ensemble Classifiers. In 2024 International Conference on Emerging Research in Computational Science (ICERCS) (pp. 1-6). IEEE.

[2]. Naseer, M., Ullah, F., Saeed, S., Algarni, F. and Zhao, Y., 2025. Explainable TabNet ensemble model for identification of obfuscated URLs with features selection to ensure secure web browsing. Scientific Reports, 15(1), p.9496.

[3]. Fatima, R., Fareed, M.M.S., Ullah, S., Ahmad, G. and Mahmood, S., 2024. An Optimized Approach for Detection and Classification of Spam Email's Using Ensemble Methods. Wireless Personal Communications, pp.1-27.

[4]. Oluchukwu, U.W., Sylvanus, O.A., Asogwa, D., Chinedu, E., Chibuogu, A. and Sylvanus, A.K., 2024. Hybrid machine learning algorithms for email and malware spam filtering: A review. Eur. J. Theor. Appl. Sci, 2, pp.76-86.

[5]. Shawly, T., Alsheikhy, A.A., Said, Y., Shaaban, S.M., Lahza, H., AbuEid, A.I. and Alzahrani, A., 2025. DaC-GANSAEBF: Divide and Conquer-Generative Adversarial Network—Squeeze and Excitation-Based Framework for Spam Email Identification. Computer Modeling in Engineering & Sciences (CMES), 142(3).

[6]. Truong, C.K., Hao Do, P. and Duc Le, T., 2023. A comparative analysis of email phishing detection methods: a deep learning perspective.

[7]. Ratmele, A., Dhanare, R. and Parte, S., 2025. Octave convolutional multi-head capsule nutcracker network with oppositional Kepler algorithm based spam email detection. Wireless Networks, 31(2), pp.1625-1644.

[8]. Rashed, A., Abdulazeem, Y., Farrag, T.A., Bamaqa, A., Almaliki, M., Badawy, M. and Elhosseini, M.A., 2025. Toward Inclusive Smart Cities: Sound-Based Vehicle Diagnostics, Emergency Signal Recognition, and Beyond. Machines, 13(4), p.258.

[9]. Crawford, M., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N. and Al Najada, H., 2015. Survey of review spam detection using machine learning techniques. Journal of Big Data, 2, pp.1-24.

[10]. Kumar, S., Kar, A.K. and Ilavarasan, P.V., 2021. Applications of text mining in services management: A systematic literature review. International Journal of Information Management Data Insights, 1(1), p.100008.

[11]. Thudumu, S., Branch, P., Jin, J. and Singh, J., 2020. A comprehensive survey of anomaly detection techniques for high dimensional big data. Journal of big data, 7, pp.1-30

